

Research article

A novel optimization approach to estimating kinetic parameters of the enzymatic hydrolysis of corn stover

Fenglei Qi and Mark Mba Wright *

Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA

* **Correspondence:** Email: markmw@iastate.edu; Tel: +1-515-294-0913.

Abstract: Enzymatic hydrolysis is an integral step in the conversion of lignocellulosic biomass to ethanol. The conversion of cellulose to fermentable sugars in the presence of inhibitors is a complex kinetic problem. In this study, we describe a novel approach to estimating the kinetic parameters underlying this process. This study employs experimental data measuring substrate and enzyme loadings, sugar and acid inhibitions for the production of glucose. Multiple objectives to minimize the difference between model predictions and experimental observations are developed and optimized by adopting multi-objective particle swarm optimization method. Model reliability is assessed by exploring likelihood profile in each parameter space. Compared to previous studies, this approach improved the prediction of sugar yields by reducing the mean squared errors by 34% for glucose and 2.7% for cellobiose, suggesting improved agreement between model predictions and the experimental data. Furthermore, kinetic parameters such as K_{2IG2} , K_{1IG} , K_{2IG} , K_{1IA} , and K_{3IA} are identified as contributors to the model non-identifiability and wide parameter confidence intervals. Model reliability analysis indicates possible ways to reduce model non-identifiability and tighten parameter confidence intervals. These results could help improve the design of lignocellulosic biorefineries by providing higher fidelity predictions of fermentable sugars under inhibitory conditions.

Keywords: Multi-objective regression; enzymatic hydrolysis kinetics; parameter estimation; multi-objective particle swarm optimization; likelihood profile

Nomenclature

A acetic acid concentration, g/kg

K_{IG} inhibition constants for glucose in each reaction, g/kg

b_i	enzyme activity decreasing factor, kg/g	K_{iIG2}	inhibition constants for cellobiose in each reaction, g/kg
C_5	C5 sugar concentration, g/kg	K_{iIX}	inhibition constants for C5 sugars in each reaction, g/kg
E_{1max}	maximum mass of CBH and EG that can adsorb onto one unit mass of substrate, 0.06 g/g	K_{3M}	cellobiose saturation constant, g/kg
E_{2max}	maximum mass of β -glucosidase that can adsorb onto one unit mass of substrate, 0.01 g/g	k_{ir}	reaction rates for each reaction, g/kg/h
E_{1B}	bound concentration of CBH and EG, (g/kg)	N	number of experimental measurement points
E_{2B}	bound concentration of β -glucosidase, (g/kg)	p	number of parameters in the model
E_{1F}	free concentration of CBH and EG, (g/kg)	S	substrate concentration, g/kg
E_{2F}	free concentration of β -glucosidase, g/kg	S_0	substrate initial concentration, g/kg
E_{iT}	total enzyme concentration, g/kg	α	quantile of the χ^2 -distribution
$f_{\beta G}$	fraction of the maximum β -glucosidase activity	df	degree of freedom
G	glucose concentration, g/kg	\mathbf{V}	objective matrix
G_2	cellobiose concentration, g/kg	\mathbf{x}	vector of state variables
K_{1ad}	dissociation constant for the CBH and EG adsorption/desorption reaction, 0.4 g/g	\mathbf{y}	vector of observables
K_{2ad}	dissociation constant for the β -glucosidase adsorption/desorption reaction, 0.1 g/g	θ	vector of model parameters

1. Introduction

Global concerns over climate impacts of greenhouse gas (GHG) emissions have led to the pursuit of low carbon intensity alternatives to fossil fuels [1]. These efforts led to the commercialization of corn grain and sugarcane based ethanol to replace petroleum derived fuels in the transportation sector [2]. However, competition with food production and the desire for greater GHG reductions prompted the development of lignocellulosic biomass conversion technologies [3,4]. The conversion of lignocellulosic biomass into ethanol is a sustainable approach to produce renewable fuels with significant reduction of GHG emissions [5]. Economic and performance challenges have limited the commercial adoption of lignocellulosic ethanol technologies [6].

Lignocellulosic biomass such as corn stover is mainly composed of cellulose, hemicellulose and lignin intertwined by a complex matrix formed by these three biopolymers. This protective matrix is a detriment to the biological conversion of lignocellulosic biomass into fermentable sugars. Therefore, scientists have developed various methods to extract sugars from lignocellulosic biomass [7]. Enzymatic hydrolysis combined with feedstock pretreatment is preferred over other chemical hydrolysis for its higher yield, minimal byproduct formation, low energy requirements, and mild operating conditions [8,9]. However, this technique faces technical challenges and the process is not yet economically feasible [10,11].

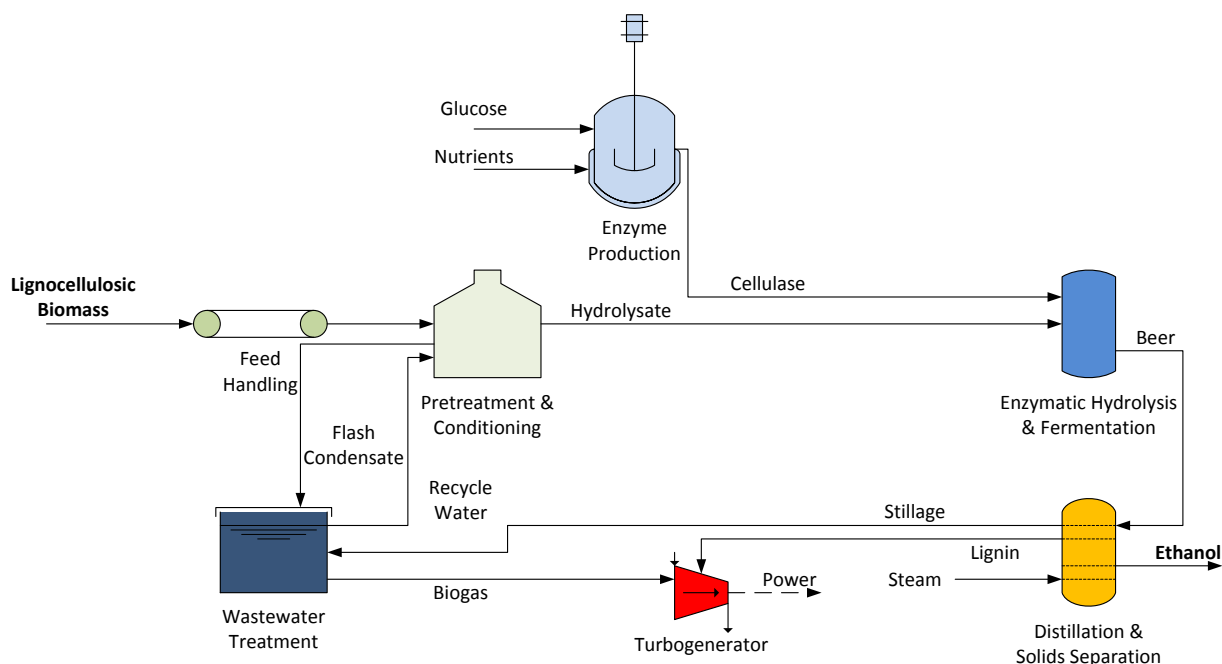


Figure 1. Lignocellulosic biomass ethanol production via enzymatic hydrolysis and fermentation.

Enzymatic hydrolysis is an important process in the conversion of lignocellulosic biomass into ethanol. As shown in Figure 1, lignocellulosic biomass is first fed into a pretreatment and conditioning process, then it proceeds to the enzymatic hydrolysis and fermentation units where it mixes with cellulase enzymes. Hydrolysis yields sugars that are fermented into ethanol. The beer mixture can be distilled into high purity ethanol. Process by-products include stillage and lignin which can be combusted to generate power and steam. The National Renewable Energy Laboratory has described the details of this process and estimated that ethanol can be produced at a minimum fuel-selling price of \$2.15 per gallon [12].

Process modeling contributes to biofuel process development through model-based evaluation of the integrated operation of hydrolysis and co-fermentation process [13] and model-based optimization of bioprocesses [14]. In particular, kinetic modeling of enzymatic hydrolysis allows for performing realistic process operation simulations which improves biorefinery design and optimization. The fidelity of the modeling significantly depends on the availability of a reliable enzymatic hydrolysis kinetic model which can reflect the main reaction rates and activities from reactants to products in the process.

Previous studies developed various kinetic models for enzymatic hydrolysis of cellulose substrate based on empirical data or fundamental mechanistic models, which have recently been reviewed in Bansal et al. [15]. Mechanistic models are often preferred to empirical models because empirical models are only applicable to the conditions under which they are developed and do not completely characterize the major physical and chemical activities in the process. Furthermore, mechanistic models provide insights into the major chemical activities of enzymes and substrates and allow simulating conditions that lie outside experimental conditions. However, developing robust mechanistic models for enzymatic processes is challenging due to limited informative experimental

datasets and the complex nature of lignocellulosic feedstocks. Therefore, recent studies focus on developing semi-mechanistic kinetic models describing major reaction activities.

One of the extensively cited semi-mechanistic kinetic models for enzymatic hydrolysis of cellulose was proposed by Kadam et al. [16]. This model incorporates parameters for enzyme adsorption, sugar inhibitions, temperature effects, and substrate reactivity. Subsequent studies have refined this model: Zheng et al. [17] considered the adsorption of enzyme onto lignin in their model; Tsai et al. [18] combined the Kadam's model [16] with a transglycosylation reaction at high glucose levels; Scott et al. [19] recently modified the model by introducing acetic acid inhibition effects and considering changes in bounded enzyme activity based on experimental observations.

Model parameter estimation from experimental data is crucial in the model development. A consensus has been reached that kinetic parameter estimation with multi-response data is favored in terms of parameter reliability [20-22]. Central to this realization in the model development is the formation of optimization objective functions from multiple responses. Although least squares or weighted least squares functions are often formulated as the optimization targets in previous model developments [16-19], they have been reported to have limitations when dealing with multi-response data [21,23]. In order to avoid this problem, multiple objectives are optimized simultaneously instead of combining them into one single objective.

Providing confidence intervals for estimated model parameters is necessary for assessing the reliability of the developed model. A profile likelihood method is adopted in this research as suggested by Raue et al. [24]. This work demonstrates a novel method for improving parameter estimates of kinetic models of enzymatic hydrolysis.

2. Hydrolysis kinetic model

The multireaction model proposed by Kadam et al.[16] is sophisticated enough to describe the complexities of enzymatic hydrolysis of lignocellulosic biomass. It includes adsorption mechanism of cellulase components onto substrate, considers end-product (C6 and C5 sugar) inhibitions and incorporates substrate reactivity change due to variations in crystal structure, degree of polymerization and substrate accessibility among other reasons. Based on experimental observations, Scott et al. [19] made some modifications about this model: elimination of the inhibitory effect of xylose on β -glucosidase, incorporation of acetic acid inhibitory effect and consideration of the change of the activity of the adsorbed enzymes. The modified reaction scheme is illustrated in Figure 2. As shown, there are three major reactions: 1) cellulose conversion to cellobiose, 2) cellulose conversion to glucose, and 3) cellobiose conversion to glucose. Dashed lines indicate inhibitory effects caused by acetic acid, sugars, and intermediate products.

Enzyme adsorption onto the lignocellulosic substrates have been described by a Langmuir model [25] formulated as Equation (1):

$$E_{iB} = \frac{E_{i\max} K_{iad} E_{iF} S}{1 + K_{iad} E_{iF}} \quad i = 1, 2. \quad (1)$$

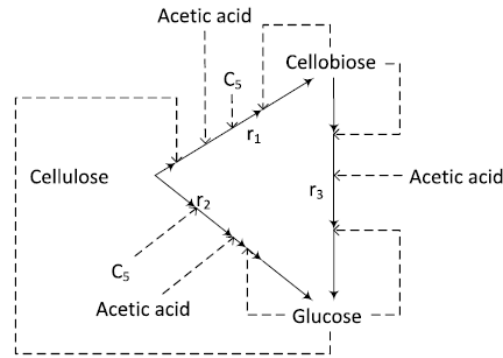


Figure 2. Reaction scheme for cellulose enzymatic hydrolysis. Enzymes involved in r_1 : endo- β -1,4-glucanase (EC 3.2.1.4) and exo- β -1,4-cellobiohydrolase (CBH) (EC 3.2.1.91). Enzymes involved in r_2 : exo- β -1,4-cellobiohydrolase (CBH) (EC 3.2.1.91) and exo- β -1,4-glucan glycohydrolase (EC 3.2.1.74). Enzymes involved in r_3 : β -glucosidase. Dashed lines show the inhibition effects of sugars and acids on enzymes (adapted from Kadam et al. [16]).

Equations (2)–(4) describe the reaction rates shown in Figure 2 are:

$$r_1 = (1 - b_1 S_0) \frac{k_{1r} E_{1B} R_S S}{1 + G_2 / K_{1G2} + G / K_{1IG} + C_5 / K_{1IX} + A / K_{1IA}} \quad (2)$$

$$r_2 = (1 - b_2 S_0) \frac{k_{2r} E_{1B} R_S S}{1 + G_2 / K_{2IG2} + G / K_{2IG} + C_5 / K_{2IX} + A / K_{2IA}} \quad (3)$$

$$r_3 = \frac{f_{\beta G} k_{3r} E_{2F} G_2}{K_{3M} (1 + G / K_{3IG} + A / K_{3IA}) + G_2} \quad (4)$$

where r_1 represents the reaction rate for the Cellulose-to-Cellobiose reaction with competitive glucose, cellobiose, C5 sugar and acetic acid inhibition, r_2 is reaction rate for the Cellulose-to-Glucose reaction with competitive glucose, cellobiose, C5 sugar and acetic acid inhibition, r_3 is the reaction rate for the Cellobiose-to-Glucose reaction with competitive glucose and acetic acid inhibition. The substrate reactivity is correlated as

$$R_S = \frac{S}{S_0} \quad (5)$$

Mass balances for cellulose, cellobiose, glucose and enzymes are

$$\frac{dS}{dt} = -r_1 - r_2 \quad (6)$$

$$\frac{dG_2}{dt} = 1.056r_1 - r_3 \quad (7)$$

$$\frac{dG}{dt} = 1.111r_2 + 1.053r_3 \quad (8)$$

$$E_{iT} = E_{iF} + E_{iB} \quad (9)$$

where, 1.056, 1.111 and 1.053 are molecular mass ratio of Cellobiose/Cellulose (substrate), Glucose/Cellulose, and Glucose/Cellobiose.

This mathematical model considers the effects of substrate and enzyme loadings, sugar and acid inhibitions, and decreasing activities of substrate and bounded enzymes on the production of glucose. In general, increasing substrate loading or decreasing enzyme loading reduces the yield of glucose. Increasing sugar and acid content impede the production of glucose. As reported in [19], the substrate loading also affects the amount of enzyme that can be bounded to substrate. In order to estimate the parameters quantifying these relationships, it is preferable to select experimental datasets that measure the production of glucose and cellobiose with respect to substrate loading, enzyme loading, sugar and acid content.

3. Methods

The general methodology adopted by this study is shown in Figure 3. First, we process experimental data and design an appropriate kinetic reaction model. Then, we calibrate the model parameters to gather initial guesses before optimizing the model objective functions. Finally, we analyze the reliability of the developed model and calculate parameter confidence levels to estimate the uncertainty ranges for the obtained parameters. An optional step is to modify the kinetic reaction model based on initial optimization results and model reliability analysis. This is often needed when the initial model is not an accurate representation of the experimental process.

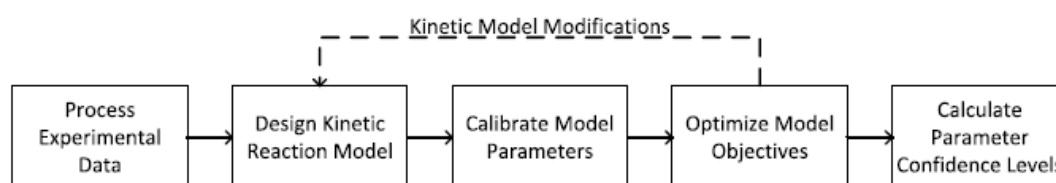


Figure 3. Kinetic model design, calibration, optimization, and parameter confidence level estimate methodology.

3.1. Experimental data

This study employs experimental data published by Scott et al. [19] and shown in Table 1. This experimental dataset includes varying initial solid loadings (10–25% w/w), and the use of the pretreatment liquor and washed solids with or without supplementation of key inhibitors. Feedstock types include full slurry (F), washed solids (W), W with acetic acid, W with a sugar stock solution, and W with only Glucose (G). The glucan content in the solids varies between 53.2 and 63.2 weight

percent. Enzyme dose in filter paper units (FPU) per gram of glucan varies between 5.4 and 25.0. The table also provides initial soluble solids content in g/kg feedstock of cellobiose, glucose, xylose, arabinose, and acetic acid. A total of 12 experimental dataset were employed to calibrate the model parameters. Both glucose and cellobiose versus time were measured in the dataset. Details about the experimental setup are provided by Scott et al. [19].

Table 1. Initial experimental conditions for enzymatic hydrolysis of cellulose to sugars (adapted from [19])

No.	Feedstock ^a	SF ^b	Glucan ^c	E_T ^d	G_{20}	G_0	X_0	Ar_0	Ac_0
Exp1	F	0.15	63.2	25.0	1.26	24.07	63.87	10.57	12.20
Exp2	F	0.15	63.2	8.2	1.26	24.17	64.20	10.57	12.20
Exp3	W	0.10	60.1	9.6	0.0	0.0	0.0	0.0	0.0
Exp4	F	0.10	60.1	9.6	1.57	16.47	44.14	7.01	8.16
Exp5	F	0.15	63.2	9.6	1.26	24.05	64.44	10.57	12.20
Exp6	W+Ac. A	0.15	63.2	9.6	0.0	0.0	0.0	0.0	12.20
Exp7	W	0.15	63.2	9.6	0.0	0.0	0.0	0.0	0.0
Exp8	W+Sug.	0.15	63.2	9.6	1.26	24.05	64.44	10.57	0.0
Exp9	W	0.13	60.1	9.6	0.0	0.0	0.0	0.0	0.0
Exp10	W+G	0.10	53.2	5.4	0.0	140.07	0.0	0.0	0.0
Exp11	W	0.25	53.2	16.2	0.0	0.0	0.0	0.0	0.0
Exp12	W	0.25	53.2	5.4	0.0	0.0	0.0	0.0	0.0

Initial soluble solids content [g/kg]: G_{20} : cellobiose; G_0 : glucose; X_0 : xylose; Ar_0 : arabinose; Ac_0 : acetic acid.

^aF: full slurry; W: washed solids; Ac. A: acetic acid; Sug: sugar stock solution; G: glucose

^bSolid fraction as weight percent of insoluble solids.

^cGlucan percentage content in solids.

^dEnzyme dose in filter paper units (FPU) per gram of glucan.

3.2. Parameter estimation

Equation (6)–(9) can be written in a general form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, t) \quad (10)$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t)) \quad (11)$$

where, \mathbf{f} is reaction rate vector composed of Equation (2)–(4). State variables $\mathbf{x}(t)$ correspond to the measured concentrations of cellulose (S), cellobiose (G_2), and glucose (G) in Equation (2)–(4). $\boldsymbol{\theta}$ is p -dimensional vector of model parameters to be estimated. \mathbf{g} is mapping function vector from state variables $\mathbf{x}(t)$ to the observables $\mathbf{y}(t)$, which is usually a conversion from measured variables to target variables.

The estimation of model parameters $\boldsymbol{\theta}$ is achieved by optimizing the objective functions which represent the fitness of the observables predicted by the model to the experimental data. The objective functions are usually written as mean squared residuals (also called mean squared error)

and mean cross-product of residuals, and constitute an objective matrix \mathbf{V} . In this calibration, there are two observables (glucose and cellobiose) from experimental measurements [19]. The objective matrix has 2 by 2 dimensions and it can be written as

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{1n}^i - y_1^i(\boldsymbol{\theta}, t_n))^2 & \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{1n}^i - y_1^i(\boldsymbol{\theta}, t_n))(y_{2n}^i - y_2^i(\boldsymbol{\theta}, t_n)) \\ \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{1n}^i - y_1^i(\boldsymbol{\theta}, t_n))(y_{2n}^i - y_2^i(\boldsymbol{\theta}, t_n)) & \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{2n}^i - y_2^i(\boldsymbol{\theta}, t_n))^2 \end{bmatrix}$$

where, y_{1n}^i denotes n^{th} data point for 1st observable measured at time point t_n in the i^{th} experiment, $y_1^i(\boldsymbol{\theta}, t_n)$ is the model prediction of 1st observable at time point t_n under the i^{th} experimental condition. In our study, the diagonal elements of matrix \mathbf{V} are chosen as the objectives. The two objective optimization problem is formulated as

$$\hat{\boldsymbol{\theta}} = \arg \min \left[\frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{1n}^i - y_1^i(\boldsymbol{\theta}, t_n))^2, \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{2n}^i - y_2^i(\boldsymbol{\theta}, t_n))^2 \right] \quad (12)$$

subject to $\boldsymbol{\theta}_{\min} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{\max}$.

The two formulated objectives indicate the difference between model predictions and experimental measurements of cellobiose concentration and glucose concentration separately. In most situations, the experimental measurement accuracy for different variables is not the same. The advantage of the proposed multi-objective regression method is to allow the user to obtain a set of solutions which have different objective combinations. The user can determine the final solutions afterwards based on their scientific or empirical judgment.

3.3. Model reliability analysis and parameter confidence intervals

Model reliability analysis is based on the likelihood profiling method as suggested in [24]. Accurate confidence intervals can also be derived from the model reliability analysis. The idea of the approach is to explore the parameter space for each parameter in the direction of the least increase of the determinant of objective matrix $|\mathbf{V}|$, it is achieved by

$$|\mathbf{V}|(\theta_k) = \arg \min_{\theta_j \neq \theta_k} \left[\frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{1n}^i - y_1^i(\boldsymbol{\theta}, t_n))^2, \frac{1}{\sum N_i} \sum_{i=1}^D \sum_{n=1}^{N_i} (y_{2n}^i - y_2^i(\boldsymbol{\theta}, t_n))^2 \right] \quad (13)$$

where, component θ_k ($k = 1, 2, \dots, p$) is fixed in a single optimization and the rest parameters are re-calibrated with the initial values from $\hat{\boldsymbol{\theta}}$. The k^{th} dimension of parameter space is gradually explored by changing θ_k . In this sense, the direction of the least increase in likelihood $|\mathbf{V}|$ can be

found after the exploration of θ_k .

The upper limit of $|\mathbf{V}|$ increase in terms of θ_k is controlled by

$$N \ln \frac{|\mathbf{V}(\boldsymbol{\theta})|}{|\mathbf{V}(\hat{\boldsymbol{\theta}})|} \leq \chi^2(\alpha, df), \quad (14)$$

where $\hat{\boldsymbol{\theta}}$ is the best estimate and N is the total number of experimental points. α is the significant level and df is the degree of freedom in the chi-squared distribution. The slope of $|\mathbf{V}|$ in the least increase direction indicates the reliability of the model. If the slope is small, it takes long steps to reach the upper limit of $|\mathbf{V}|$, so that the parameter has a wide confidence interval. Otherwise, the parameter has a narrow confidence interval.

3.4. Implementation

Multi-objective parameter estimation is achieved by adopting a novel approach based on a Multi-Objective Particle Swarm Optimization (MOPSO) method [26]. MOPSO is a novel technique designed to address global optimization problems with multiple, competing objectives. This class of problems yield multiple possible solutions of equal merit known as a Pareto optima. Additionally, the solution sets provide statistical information regarding the confidence levels of the parameters. Details regarding our MOPSO implementation are provided in the Supplementary.

The code is written in C++ and Object-Oriented Programming concept is applied. The basic structure of the algorithm is shown in Figure 4. A data unit class is employed in order to enhance data transfer between other three classes: ExpData, Reaction and Residual. ExpData mainly composes experimental data input and data elements. Reaction includes reactor model implementation and an ordinary differential equation (ODE) solver wrapper. CVODE library [27] is used to integrate ODEs

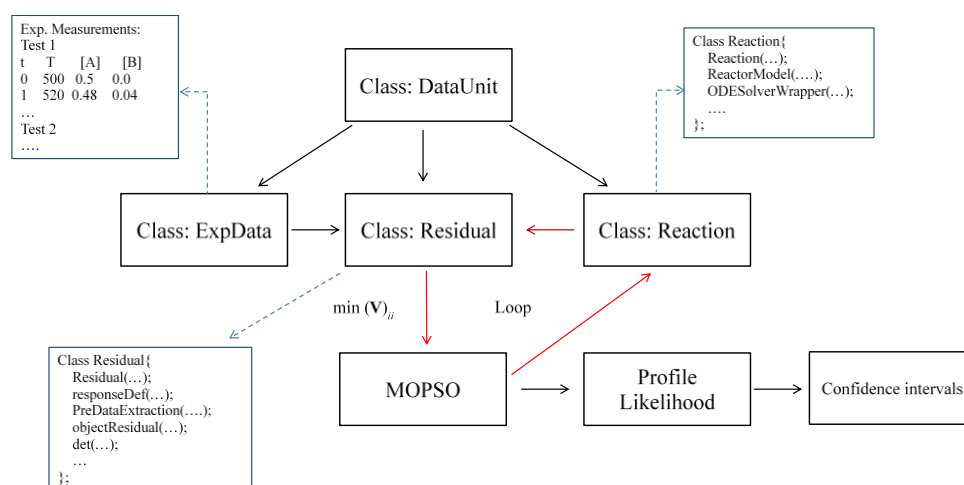


Figure 4. Schematic of algorithm structure.

in this study. Class Residual implements mapping from state variables to response variables, extracting prediction data to match with experimental dataset, and calculating objective values.

4. Results and Discussion

Parameter calibration was accomplished by selecting initial values identified in the analysis of Scott et al. [19]. These values were determined by a weighted least squares regression of the model described in Figure 2 to the experimental data. An additional parameter estimation searched through a wide parameter range that is 100 times larger or smaller than the initial parameter values. The optimization model goal was to minimize the residual errors between the kinetic model predictions and experimental data. For this purpose, we developed a multi-objective formulation of the kinetic model that simultaneously minimizes residual errors between predicted and observed yields of glucose and cellobiose. This is an important feature of the algorithm because it allows the user to determine which objective is more important and select optimal parameters based on multiple criteria.

4.1. Comparison of kinetic model predictions to experimental data

This study compares the performance of the kinetic model parameters gathered by the modified MOPSO algorithm to Scott et al. [19]. Figure 5 compares the mean squared errors of cellobiose and glucose between the two models. The MOPSO algorithm generates a set of solutions and each solution is an estimate of model parameters. The solution sets constitute a Pareto front in the corresponding objective function coordinates. Compared to the best estimate obtained by Scott et al. [19], some MOPSO solutions improve on a single objective while the other objective is degraded. A subset of MOPSO solutions achieve lower mean squared errors for both cellobiose and glucose in the lower left corner.

The minimum mean squared error limit for the first objective (glucose) is approximately 6.6, and for the second objective (cellobiose) the limit is approximately 2.3. The Pareto front shows solutions for which reducing the error in one objective can only be accomplished by increasing the error in the other objective. The parameters gathered by Scott et al. [19] achieve residual errors of 12.69 and 2.98 for glucose and cellobiose (see Table 2), respectively. Although all the Pareto solutions are optimal, solutions in the lower left (marked as black circular dots) of Figure 5 are attractive because of their improved fitness for both glucose and cellobiose yields.

Table 2 shows two selected solutions from MOPSO and Scott et al. [19]. The mean squared errors for glucose yield predictions decreased from 12.69 to 8.317 for MOPSO1—a 34% reduction; for cellobiose the error reduction was only 2.7 % and even increased in the other case. According to the determinant criteria (Equation 14), MOPSO estimates (both MOPSO1 and MOPSO2) achieve a statistically significant improvement compared to the previous model. Furthermore, some parameters such as K_{IG2} , K_{IIX} , and k_{1r} are orders-of-magnitude different than the original values although they achieve similar predictions. $f_{\beta\text{G}}$ is kept to 1.0 since there is no evidence showing the decreasing activity of β -glucosidase enzyme.

Table 2. Parameter estimates from MOPSO and Scott et al. [19].

	K_{11G2}	K_{21G2}	K_{11X}	K_{21X}	k_{1r}	K_{11G}	k_{2r}	K_{21G}	k_{3r}	K_{3M}
MOPSO1	6.28e-4	63.1050	6.843e-3	4.5515	2451.23	877.9974	4.894e-2	20120.4	20590.4	1.9861
MOPSO2	2.5557	102.0897	18.7779	4.2967	0.8022	448.4742	4.752e-2	16264.4	17461.8	1.9106
Scott	0.041	4.264	0.395	5.877	28.65	22.658	0.422	1.00e06	128.4	0.301
	K_{31G}	K_{11A}	K_{21A}	K_{31A}	b_1	b_2	$f_{\beta G}$	R_G	R_{G2}	Det
MOPSO1	1.689e-2	10.2570	1.2223	1143.372	6.7369	4.280	1.0(f)	8.317	2.90	21.24
MOPSO2	1.967e-2	65.4988	0.5055	751.1735	6.8027	4.3254	1.0(f)	8.221	2.99	21.82
Scott	0.612	1.00e6	2.97	1.00e6	6.822	6.376	0.766	12.69	2.98	37.18

(f) represents the parameter is fixed; R_G is mean squared errors for glucose; R_{G2} is mean squared errors for cellobiose; Det represents the determinant of residual matrix.

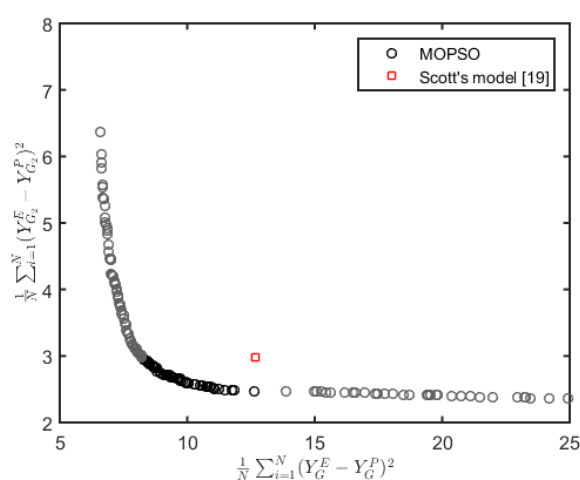


Figure 5. Kinetic model comparison of mean squared errors between model prediction and experimental data. X axis shows mean squared error of glucose and Y axis represents mean squared error of cellobiose; Y_G^E and Y_G^P are glucose concentrations measured in experiments and predicted by model; Y_{G2}^E and Y_{G2}^P are cellobiose concentrations measured in experiments and predicted by model; N is the total number of experimental points; black circular dots show the mean squared errors of glucose and cellobiose that are both smaller than Scott's model [19].

Figure 6 compares experimental data to model predictions from MOPSO and Scott et al. [19]. In this comparison, we selected the MOPSO2 solution and the estimated parameter values are shown in Table 2. As shown, both models achieve a qualitatively suitable representation of the experimental trends in all cases. The main differences observed in the predictions of glucose production are found in the 10th, 11th and 12th experimental conditions shown in Figure 6. MOPSO improves upon the Scott's model by reducing the mean squared errors of glucose from 156.76 to 48.57 in 10th experiment. The model by Scott et al. appears to overpredict the production of glucose under initial high glucose concentration, while MOPSO yields a more accurate prediction of the inhibitory effects of glucose. MOPSO overpredicts the production of glucose at lower enzyme loading as in 12th

experiment condition. However, additional experimental data would be required to validate these observations given the likelihood of uncertainty in the experimental measurements.

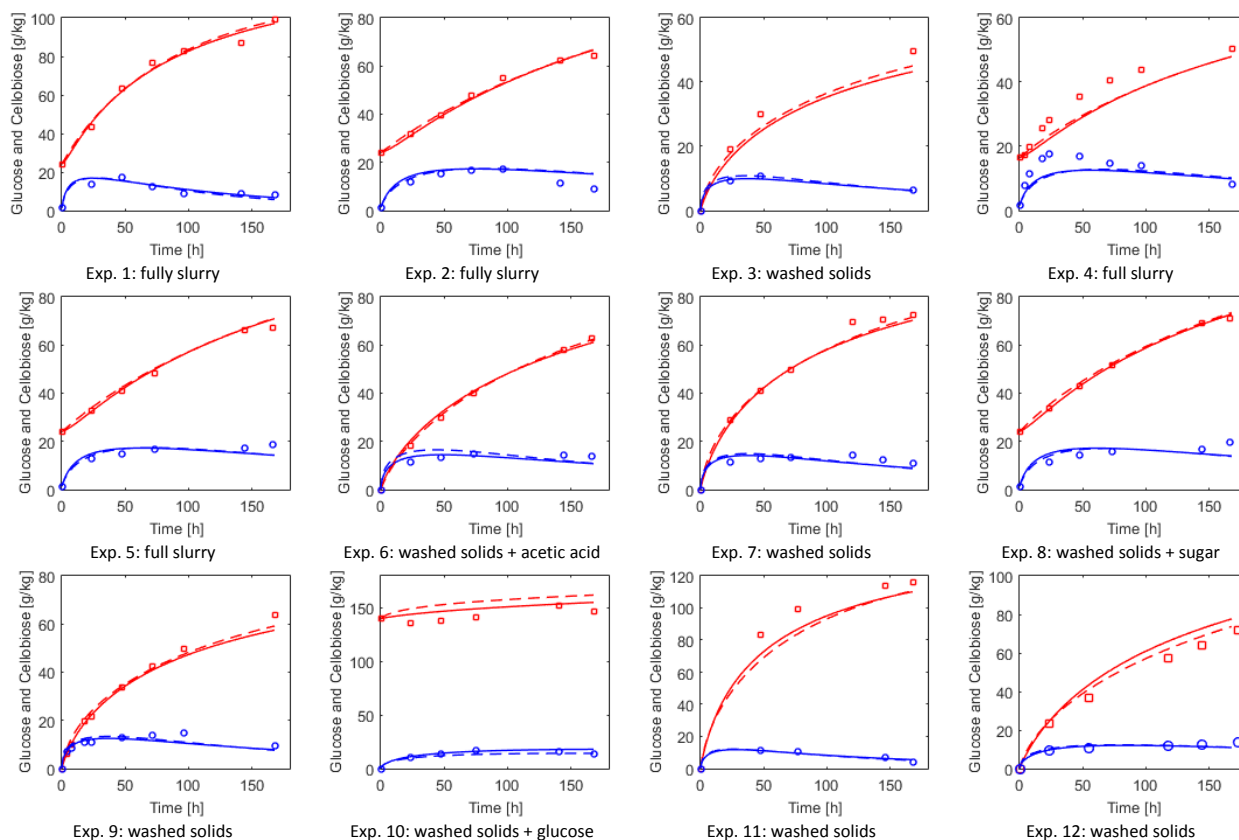


Figure 6. Experimental data and model predictions for glucose and cellobiose concentrations vs reaction time. Square dots are experimental measurements of glucose and circle dot are experimental measurements of cellobiose; Solid lines are predictions based on MOPSO estimates; Dash lines are predictions from Scott et al. [19].

4.2. Kinetic model reliability analysis and parameter confidence intervals

Figure 7 shows the result of exploiting profile likelihood along each parameter space. The profile of the likelihood in terms of parameter value provides the information of model reliability. The model reliability is reduced in the case that the likelihood profile does not increase or only increase slowly in either one side exploring direction or both sides (increase and decrease of parameter value) in the parameter coordinate. Correspondingly, the parameter has a wide confidence interval in this case. The model is usually said non-identifiable if the increase in the likelihood is slow and statistically insignificant when the parameter value increases or decreases. One of the advantages of likelihood profiling method is to allow us to visually evaluate the likelihood profile in each parameter space.

The likelihood ($\det(\mathbf{V})$) does not increase in the parameter space of K_{2IG2} , K_{1IG} , K_{2IG} , K_{1IA} , and K_{3IA} except approaching the natural lower bound of 0. Scott et al. [19] points out that reducing the number of estimated parameters can tighten the confidence intervals by fixing some parameters in

the calibration. In their best model, the above five parameters are fixed along with K_{11G2} , k_{3r} , K_{3M} , K_{31G} . The underlying reason is that these parameters are non-identifiable which results in the wide range of parameter confidence intervals when they are included in the model. Reducing the parameter redundancy as shown in Scott et al. [19] is one way to increase model reliability. An alternative is to improve experimental design. Considering the experimental design as shown in Table 1, the amount of sugars (cellobiose and glucose) and acetic acid do not change much from experiment to experiment, which is not favorable to quantifying the inhibitory effects of the sugars and acetic acid. Further design of experiments can be based on this knowledge and improve model parameter reliability.

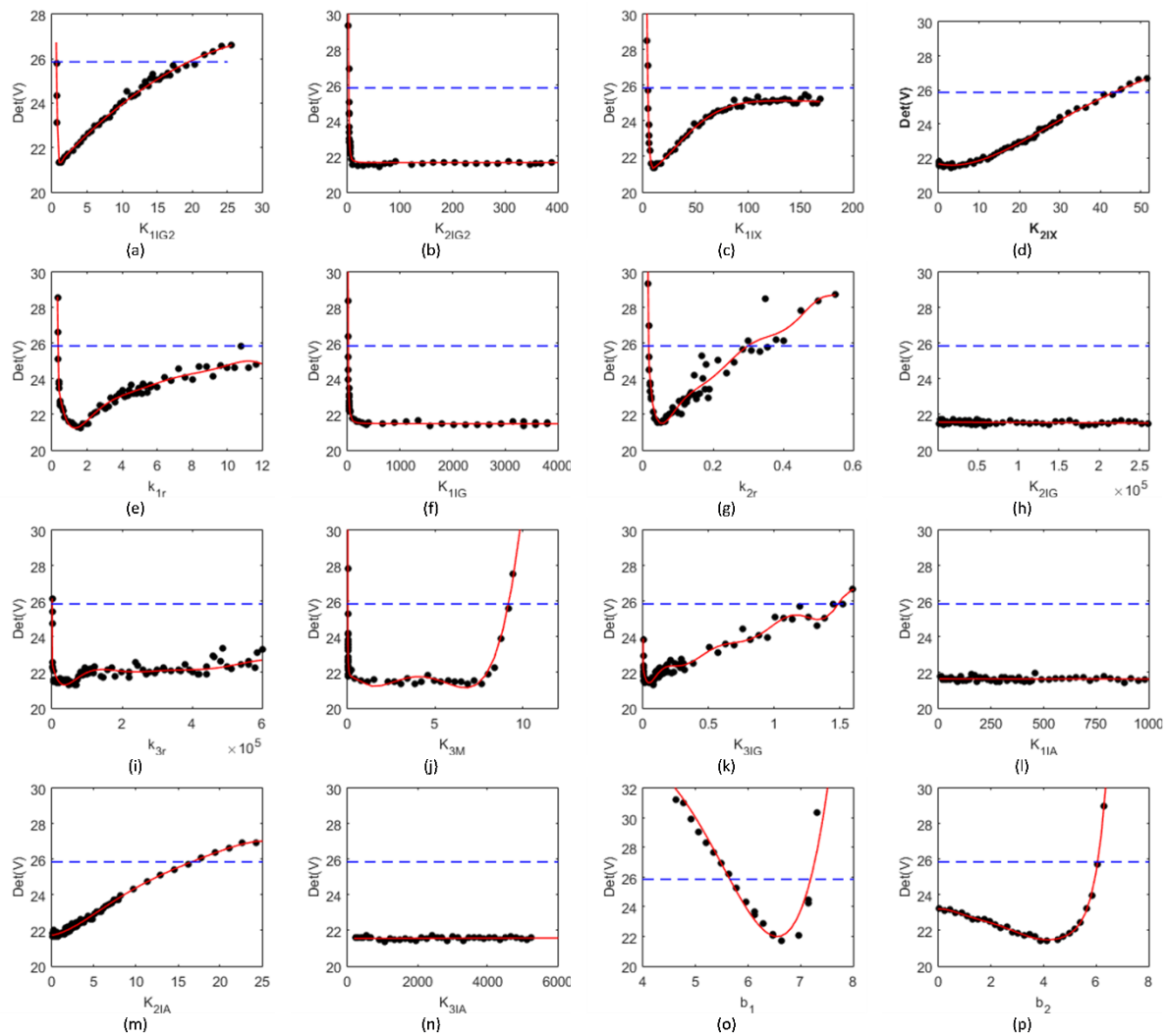


Figure 7. Exploration of profile likelihood along each parameter space. Black dots show the $\det(V)$ value achieved after re-optimization vs each parameter; solid lines are fitted trend lines; dash lines are the upper limit of $\det(V)$ according to Equation (14) with 5% percentile of a Chi-squared distribution and 16 degrees of freedom.

The likelihood in some parameters such as K_{11X} , k_{1r} , k_{3r} , K_{31G} and b_2 changes gradually but the gradient is relatively small which results in either unclosed confidence intervals or large confidence

intervals. Among all the reasons, one of them is still related to inadequate experiment design which gives good explanations for likelihood profile in terms of K_{1IX} and K_{3IG} . Correlations between parameters might be another reason, for example, k_{1r} might have strong correlations with K_{1IG} , K_{1IA} and K_{1IX} . In summary, exploiting profile likelihood is helpful to analyzing the reliability of the estimates and serve as a powerful tool in model-based experiment design.

Another virtue of exploiting likelihood profile is to obtain accurate confidence intervals. Table 3 shows upper and lower limits for the MOPSO kinetic model parameters with a 95% confidence level. Most of the parameters have wide or undefined confidence intervals, which indicates that there is insufficient experimental data to provide a robust estimate of the kinetic parameters in the model.

Table 3. Estimated MOPSO kinetic model parameter values and minimum and maximum confidence intervals at the 95% level.

	K_{1IG2}	K_{2IG2}	K_{1IX}	K_{2IX}	k_{1r}	K_{1IG}	k_{2r}	K_{2IG}
MOPSO2	2.5557	102.0897	18.7779	4.2967	0.8022	448.4742	4.752e-2	16264.4
CI ^{95%}	0.6389	2.317	4.5632	-	0.3486	14.5754	1.59e-2	-
	21.50	-	-	43.8778	-	-	0.3	-
	k_{3r}	K_{3M}	K_{3IG}	K_{1IA}	K_{2IA}	K_{3IA}	b_1	b_2
MOPSO2	17461.8	1.9106	1.967e-2	65.4988	0.5055	751.1735	6.8027	4.3254
CI ^{95%}	460.1	1.92e-2	-	-	-	-	5.7260	-
	-	9.2097	1.4476	-	16.7698	-	7.1884	6.0642

CI^{95%} refers to the confidence interval with 0.95 confidence level. (-) in the tables indicates the upper or lower intervals are undetermined based on the likelihood profile method within approximately 20 times or 1/20th times of the estimates.

5. Conclusions

Lignocellulosic ethanol production is a sustainable alternative for the production of renewable fuels. This study investigated the enzymatic hydrolysis of lignocellulosic biomass to produce fermentable sugars. Kinetic models for this process are an important part of estimating the technical and economic performance of lignocellulosic ethanol biorefineries.

This paper describes the use of a novel multi-objective parameter estimation method for developing reliable hydrolysis kinetic models. Estimates from the multi-objective regression shows a statically significant improved fit to the experimental data compared to previous studies. We achieved improved predictions for the yields of glucose and cellobiose from cellulose in the presence of high glucose content. Furthermore, we analyzed model reliability by adopting the likelihood profiling method. This method allows us to efficiently analyze which parameters contribute to model non-identifiability and identify possible ways to improve model reliability. Parameter confidence intervals are accurately determined with this method.

Comparisons of the kinetic models to experimental data indicate qualitative agreement between predicted and measured glucose and cellobiose yields. We developed inferences about the underlying phenomena and acknowledge uncertainty in the experimental measurement. Additional experimental data could improve the accuracy of the model parameter estimates and our ability to predict the performance of enzymatic hydrolysis in future research.

Acknowledgements

The authors would like to acknowledge financial support from the National Science Foundation EPSCoR program (Grant EPS-1101284) and Bioeconomy Institute of Iowa State University.

Conflict of interest

All the authors of this paper declare no conflicts of interest.

References

1. Pachauri RK, Allen M, Barros V, et al. (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
2. MacDonald T, Yowell G, McCormack M (2001) US ethanol industry: production capacity outlook. California Energy Commission. P600-01-017.
3. Kazi FK, Fortman J, Anex R, et al. (2010) Techno-economic analysis of biochemical scenarios for production of cellulosic ethanol: Citeseer.
4. Kazi FK, Fortman JA, Anex RP, et al. (2010) Techno-economic comparison of process technologies for biochemical ethanol production from corn stover. *Fuel* 89: S20-S28.
5. Winchester N, Reilly JM (2015) The feasibility, costs, and environmental implications of large-scale biomass energy. *Energy Economics* 51: 188-203.
6. Brown TR, Brown RC (2013) A review of cellulosic biofuel commercial-scale projects in the United States. *Biofuels, bioprod bioref* 7: 235-245.
7. Chiamonti D, Prussi M, Ferrero S, et al. (2012) Review of pretreatment processes for lignocellulosic ethanol production, and development of an innovative method. *Biomass Bioenerg* 46: 25-35.
8. Sun Y, Cheng J (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour technol* 83: 1-11.
9. Wald S, Wilke CR, Blanch HW (1984) Kinetics of the enzymatic hydrolysis of cellulose. *Biotechnol bioeng* 26: 221-230.
10. Hayes DJ (2009) An examination of biorefining processes, catalysts and challenges. *Catal Today* 145: 138-151.
11. Sin G, Meyer AS, Gernaey KV (2010) Assessing reliability of cellulose hydrolysis models to support biofuel process design—identifiability and uncertainty analysis. *Comput chem eng* 34: 1385-1392.
12. Humbird D, Aden A (2009) Biochemical production of ethanol from corn stover: 2008 state of technology model: Citeseer.
13. Morales-Rodriguez R, Meyer AS, Gernaey KV, et al. (2011) Dynamic model-based evaluation of process configurations for integrated operation of hydrolysis and co-fermentation for bioethanol production from lignocellulose. *Bioresour Technol* 102: 1174-1184.
14. Morales-Rodriguez R, Meyer AS, Gernaey KV, et al. (2012) A framework for model-based optimization of bioprocesses under uncertainty: Lignocellulosic ethanol production case. *Comput Chem Eng* 42: 115-129.

15. Bansal P, Hall M, Realff MJ, et al. (2009) Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnol adv* 27: 833-848.
16. Kadam KL, Rydholm EC, McMillan JD (2004) Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. *Biotechnol prog* 20: 698-705.
17. Zheng Y, Pan Z, Zhang R, et al. (2009) Kinetic modeling for enzymatic hydrolysis of pretreated creeping wild ryegrass. *Biotechnol bioeng* 102: 1558-1569.
18. Tsai CT, Morales-Rodriguez R, Sin G, et al. (2014) A dynamic model for cellulosic biomass hydrolysis: A comprehensive analysis and validation of hydrolysis and product inhibition mechanisms. *Appl biochem biotech* 172: 2815-2837.
19. Scott F, Li M, Williams DL, et al. (2015) Corn stover semi-mechanistic enzymatic hydrolysis model with tight parameter confidence intervals for model-based process design and optimization. *Bioresource technol* 177: 255-265.
20. Box GE, Draper NR (1965) The Bayesian estimation of common parameters from several responses. *Biometrika* 52: 355-365.
21. Stewart WE, Caracotsios M, Sørensen JP (1992) Parameter estimation from multiresponse data. *AIChE J* 38: 641-650.
22. Ziegel ER, Gorman JW (1980) Kinetic modelling with multiresponse data. *Technometrics* 22: 139-151.
23. Conceição EL, Portugal AA (2012) Comparison of Two Robust Alternatives to the Box–Draper Determinant Criterion in Multiresponse Kinetic Parameter Estimation. *Ind Eng Chem Res* 51: 1118-1130.
24. Raue A, Kreutz C, Maiwald T, et al. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25: 1923-1929.
25. Kim DW, Kim TS, Jeong YK, et al. (1992) Adsorption kinetics and behaviors of cellulase components on microcrystalline cellulose. *J Ferment Bioeng* 73: 461-466.
26. Coello CAC, Pulido GT, Lechuga MS (2004) Handling multiple objectives with particle swarm optimization. *Evolutionary Computation, IEEE Transactions on* 8: 256-279.
27. Hindmarsh AC, Brown PN, Grant KE, et al. (2005) SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM T Math Software* 31: 363-396.



AIMS Press

© 2016 Mark Mba Wright et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)